Eur Respir J 2002; 20: Suppl. 36, 54s–65s
DOI: 10.1183/09031936.02.00400702

European Respiratory Journal
ISSN 0904-1850

Molecular epidemiology of tuberculosis

M.V. Burgos, A.S. Pym

Molecular epidemiology of tuberculosis. M.V. Burgos, A.S. Pym. ©ERS Journals Ltd 2002.

ABSTRACT: Despite the almost 50 yrs since the introduction of curative antituberculosis drugs, *Mycobacterium tuberculosis* continues to exert an enormous toll on world health, and tuberculosis remains the world's leading cause of death due to a single infectious agent. This has stimulated research efforts into finding new tools to tackle the continuing tuberculosis pandemic.

One of the few successes to date has been the development of a new discipline, molecular epidemiology. This has added a further dimension to the classical epidemiology of tuberculosis and enhanced understanding of how *M. tuberculosis* continues to be successfully transmitted within populations. In the process, inadequacies in tuberculosis control programmes have been identified, helping accumulate resources for their improvement.

Other technologies, based on knowledge of the complete genome sequence of *M. tuberculosis*, which will provide newer tools for probing the epidemiology of tuberculosis, are now emerging. In spite of these advances, tuberculosis continues to remain a devastating infectious disease, disproportionately impacting on the world's poorest countries.

The future challenge for molecular epidemiology is to provide better understanding of the transmission dynamics of tuberculosis in these settings and to stimulate the implementation of control measures on a more global scale.

Eur Respir J 2002; 20: Suppl. 36, 54s–65s.

Division of Infectious Diseases & Geographic Medicine, Dept of Medicine, Stanford University Medical Centre, Stanford, CA, USA.

Correspondence: M.V. Burgos, Division of Infectious Diseases & Geographic Medicine, Stanford University Medical Center, Room S-156, 300 Pasteur Drive, Stanford, CA 94305-5107 USA.
Fax: 1 6504987011
E-mail: marcos@molepi.stanford.edu

Keywords: Drug resistance
fitness
genomics
molecular epidemiology
Mycobacterium tuberculosis
restriction fragment length polymorphism

Received: February 14 2002
Accepted after revision: March 13 2002

Mycobacterium tuberculosis is one of the most successful bacterial pathogens in the history of mankind. Despite antituberculosis drugs having been available for almost 50 years, *M. tuberculosis* continues to exert an enormous toll on world health (fig. 1a). Between a third and a half of the world's population is infected with *M. tuberculosis*. Each year, there are ~2 million deaths due to tuberculosis, making tuberculosis the world's leading cause of mortality due to a single infectious agent [2]. Tuberculosis is the number one cause of death among human immunodeficiency virus (HIV)-infected individuals (fig. 1b) [2]. In 1999, there were 8.4 million cases reported worldwide, and, for the year 2005, the World Health Organization projects an incidence of 10.2 million new cases [1]. This increased incidence will occur mostly in countries in Africa and Asia, where the highest prevalence of coinfection with HIV and *M. tuberculosis* occurs. The economic impact of this pathogenic synergy is particularly great because HIV disproportionately affects persons during the most productive years of their lives.

The resurgence of tuberculosis around the world has renewed interest in understanding the epidemiology and pathogenesis of this disease. One important advance in the field of tuberculosis research has been the development of molecular techniques that allow the identification and tracking of individual strains of *M. tuberculosis*. This new discipline, the molecular epidemiology of tuberculosis, began with the identification of IS*6110*, a novel mycobacterial insertion sequence which formed the basis of a reproducible genotyping technique for *M. tuberculosis*. This method is now firmly established, but is still expensive, labour-intensive and only applicable using viable culture material. Although other typing methods, at varying stages of development, appear to offer certain advantages in terms of reproducibility, cost, ease of execution and general applicability to clinical settings, IS*6110*-based typing remains the internationally accepted standard and continues to provide new insights into the epidemiology of *M. tuberculosis*.

A newer research approach, initiated by determination of the complete genome sequence of *M. tuberculosis*, is to combine these conventional molecular epidemiological techniques with developments in mycobacterial genomics [3]. The goal is to employ the array of typing techniques now available for the identification of individual strains or clonal groupings of strains with specific phenotypic characteristics, such as transmissibility, antigenicity or resistance to antimicrobial agents. These strains can then be subjected to genome-wide analysis, using techniques such as microarrays for expression profiling or detection of genomic deletions, to determine the genetic basis of these important phenotypic traits. This multidisciplinary approach could lead to important advances in understanding the pathogenesis of

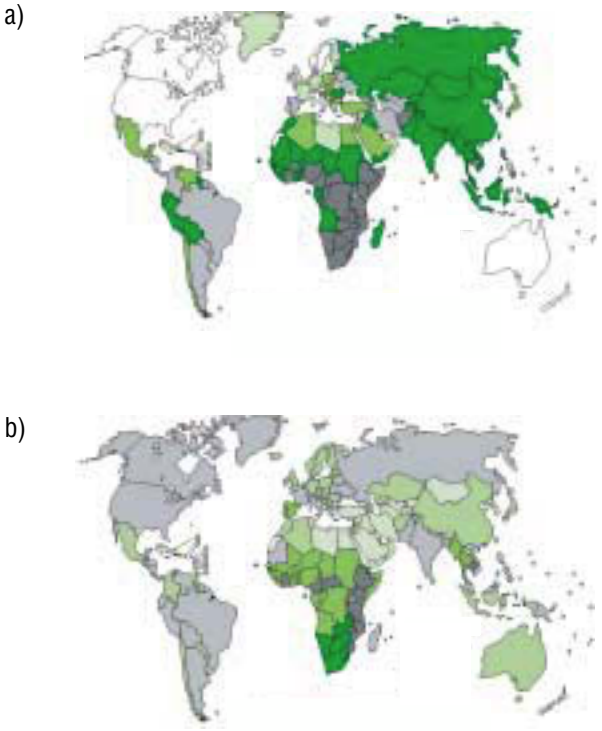

Fig. 1. – a) Estimated worldwide incidence of tuberculosis (TB) per 100,000 population. □: <10; ▒: 10–24; ▓: 25–49; ▒: 50–99; ▓: 100–300; ▓: >300. b) Estimated worldwide incidence of human immunodeficiency virus (HIV)-positive TB per 100,100 population. ▒: <0.1; ▒: 0.1–0.9; ▒: 1–9.9; ▓: 10–99; ▓: 100–249; ▓: >249; □: no data. The data are for 1999 and are an estimate based on the reported incidence of TB and HIV for each country. The incidence of TB has been rising sharply in African countries with the spread of HIV, and in Eastern Europe following the break-up of the former USSR. There have also been some successes in TB control, which have probably reduced incidence, notably in Peru and China, in comparison with previous years. Modified from [1].

human tuberculosis as well as mechanisms of drug resistance.

Molecular epidemiological markers

Before the introduction of molecular typing methods, there was little to aid distinction between individual strains of *M. tuberculosis*. Drug susceptibility patterns had been used but were of limited utility because patterns can change and resistance is rare in most populations. Typing by variability in susceptibility to infection with mycobacterial phages has also been used, but was found to be difficult to reproduce and limited by the number of phages available. The development and application of molecular typing methods, such as restriction fragment length polymorphism (RFLP) analysis based on the insertion sequence IS*6110*, in the 1990s has brought a new dimension to the study of tuberculosis and with it a new appreciation of the ecological complexities that classical epidemiology could not provide. IS*6110* is the most extensively studied and widely used of the known insertion sequences, although the availability of the *M. tuberculosis* genome sequence has led to the identification of >30 additional repetitive elements which promise to be useful typing markers [4]. This article reviews the most commonly used typing methods, newer approaches and the application of these techniques in the study of tuberculosis.

*IS*6110 *restriction fragment length polymorphism*

IS*6110* was initially described in 1989 [5]. This marker comprises 1,355 base pairs and belongs to the IS*3* family of insertion sequences. IS*6110* has two open reading frames encoding proteins required for transposition. The number of copies of IS*6110* ranges 0–25, and their positions in the *M. tuberculosis* chromosome are highly variable between different isolates (fig. 2) [7]. This variability is sufficient to generate RFLP and for it to be used in fingerprinting. IS*6110* is exclusively present in the *M. tuberculosis*

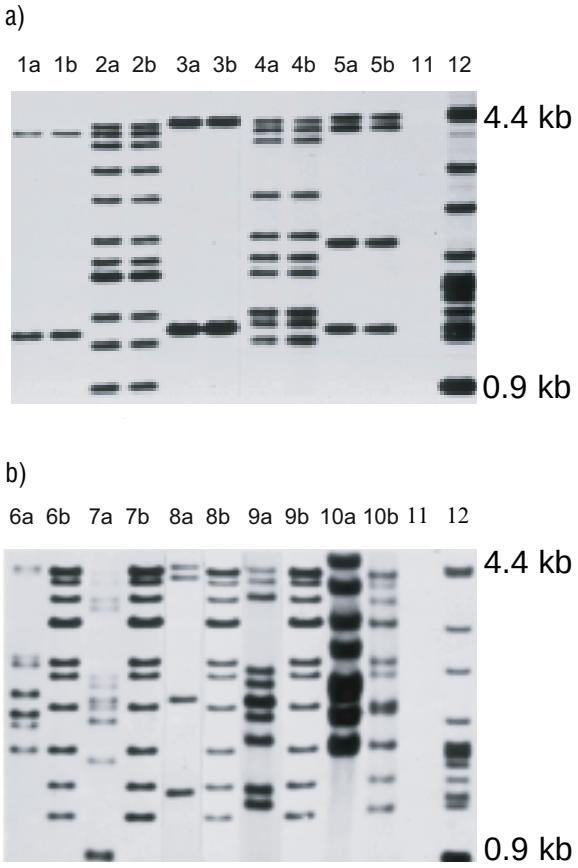

Fig. 2. – IS*6110*-based deoxyribonucleic acid fingerprinting of pairs of *Mycobacterium tuberculosis* isolates from 10 patients with two episodes of tuberculous disease. Lanes 1a–10a: strains isolated during the first episode of tuberculosis; lanes 1b–10b: isolates from the second episode; lane 11: *M. avium* (a mycobacterial species with no copies of the insertion sequence IS*6110*); lane 12: H37Rv (*M. tuberculosis* reference strain). The sizes of standard markers are indicated. a) Pairs of isolates with identical restriction fragment length polymorphisms (RFLPs), indicating that these patients experienced endogenous reactivation (relapsed) with the same strain as in their first episode of disease. b) Pairs with non-identical RFLPs, indicating exogenous reinfection with a different strain. kb: kilobase. Modified from [6].

complex species, although strains of *M. tuberculosis* lacking this element have also been described [8–11]. IS*6110* sequences are nonrandomly distributed, suggesting insertional hot spots. One of these is an area flanked by a direct repeat (DR) sequence [12]. This nonrandom distribution of IS*6110* within the genome is now recognised as a limitation to the discriminative power of typing based on the copy number and position of this sequence.

IS*6110*-based typing is the most widely applied genotyping method in the molecular epidemiology of *M. tuberculosis* and is the gold standard to which other methods are currently compared [13]. Various important factors have been identified for the standardisation of M. *tuberculosis* genetic fingerprinting, including the use of the restriction enzyme *Proteus vulgaris* II for deoxyribonucleic acid (DNA) digestion and the incorporation of molecular weight standards for the estimation of band sizes. This standardisation has facilitated the comparison of fingerprints obtained in different laboratories around the world, allowing the global dispersal of strains to be tracked. For example a HIV-seropositive patient developed primary tuberculosis from a multidrug-resistant tuberculosis (MDR-TB) strain in San Francisco, CA, USA, which was found to be a unique fingerprint in the local database. On further questioning, the patient revealed that he had been hospitalised in Buenos Aires, Argentina, the previous year, during the time of a reported large outbreak of MDR-TB [14]. The San Francisco strain was found to have an identical IS*6110* RFLP pattern and antibiotic resistance profile to the Buenos Aires strains, demonstrating the geographical extent to which individual strains of *M. tuberculosis* can be disseminated. A similar situation occurred with a HIV-seropositive patient that developed tuberculosis in Holland who was shown to have acquired tuberculosis after exposure in a hospital in Spain [15]. The standardised methodology and centralised databases made it possible to confirm that these patients acquired their infections while abroad and not in the area in which they lived.

Although the number of copies of IS*6110* can range 0–25, population-based molecular epidemiological studies report that most strains contain 8–18 copies, a number sufficient to enable discrimination between the majority of strains. For example data from San Francisco show that, of 1,800 patients over a period of 9 yrs, 1,117 (62.2%) had distinct DNA fingerprints and 683 (37.9%) were in 171 clusters sharing identical patterns. Less than 1% of the strains identified had two or less bands [11]. However, there are geographical areas in Asia and Africa in which the diversity of IS*6110* is considerably reduced [8–10, 16, 17]. In addition, greater numbers of IS*6110* bandless strains have been detected in Asia [7, 8]. The lack of polymorphism associated with low copy numbers limits the discriminatory power and the epidemiological inferences that can be drawn with this typing method. However, additional or secondary typing systems can be used to discriminate between strains with few copies [10, 18–22].

An important characteristic of a genetic marker is its stability over time. Markers which change too rapidly obscure epidemiological links, whereas those that are too stable infer direct links where they do not exist. Understanding the rate and determinants of IS*6110* pattern change is therefore important for optimal interpretation of fingerprint patterns. One study found that the banding patterns of isolates collected $\geqslant 90$ days apart from the same patient were different in 29% of cases [23]. In the Netherlands, the half-life of change in IS*6110* was estimated to be 3.2 yrs [24]. Interestingly, extensive serial passage of strains *in vitro*, such as for bacille Calmette-Guérin vaccine (attenuated tuberculosis vaccine) or of *M. tuberculosis* strains, is not associated with such rapid change in RFLP fingerprint profiles [10, 25]. It is possible, therefore, that either the studies involving serial secretors overestimate the rate of change or that replication within a host can provoke IS*6110* transposition.

Secondary markers

Empirical evidence demonstrates that the certainty with which epidemiological links can be inferred between patients infected with *M. tuberculosis* is markedly reduced if the strains involved yield less than five IS*6110* hybridizing bands, but can be improved using an additional marker [18–20, 26–29]. Thus genotyping with a secondary marker is essential for determining whether transmission has occurred in these cases. Two techniques, polymorphic guanine-cytosine-rich repetitive sequence (PGRS) profiling and spoligotyping, have been the most extensively used for this purpose.

Polymorphic guanine-cytosine-rich repetitive sequence restriction fragment length polymorphism typing

The identification of a PGRS present in multiple chromosomal clusters [30–32] enabled the development of a second RFLP typing system. This has been shown to have a discriminatory power close to that of IS*6110* typing, even in isolates with low copy numbers of IS*6110*, making it an ideal secondary typing system [21, 26, 27, 33]. However, the PGRS regions comprise many nonperfect repeats, making the RFLP patterns complex and sometimes difficult to interpret. Initially, these repetitive sequences were thought to be noncoding, but analysis of the *M. tuberculosis* genome has established that they code for the C-termini of a novel class of proteins of unknown function.

Spoligotyping

The use of DNA RFLP analysis to distinguish between strains of *M. tuberculosis* is hindered by the need to culture this organism. IS*6110* RFLP analysis is also technically demanding and costly, requiring expensive computer software for interpretation and comparison of fingerprints. Conversely, polymerase chain reaction (PCR)-based methods require very small amounts of DNA, which can be obtained without resorting to bacterial culture. In addition,

PCR can be carried out directly on clinical specimens, enabling simultaneous identification and strain typing of *M. tuberculosis* in sputum from patients [34–44].

Spoligotyping, which interrogates a DR sequence comprising a repetitive 36-base-pair element separated by short nonrepetitive sequences, is one such PCR-based technique [30, 39]. Using one set of primers, it is possible to simultaneously amplify all of the unique nonrepetitive sequences, or spacers, between the direct repeats. The presence or absence of spacers is then determined *via* Southern hybridisation. Individual strains are distinguished by the number of spacers missing from the complete spacer set defined by sequencing this region from a large number of *M. tuberculosis* strains.

Spoligotyping has been shown to be helpful when discriminating between isolates of *M. tuberculosis* with few IS*6110* bands [40, 42]. Atypical strains of mycobacteria have been analysed using spoligotyping and do not give a signal, indicating the specificity of this technique for *M. tuberculosis* [42]. The ability to perform spoligotyping directly on sputum samples makes it applicable in acute clinical situations [39, 45]. Another advantage of this secondary marker technique is that it is economical, easy to perform and a rapid means of typing the *M. tuberculosis* complex. These characteristics make it a candidate for use in resource-poor situations. Nonetheless, the discriminatory power of this method is less than that of IS*6110* typing [40, 46–48]. Since *M. tuberculosis* isolates with a different spoligotype invariably have distinguishable IS*6110* profiles, spoligotyping could conceivably be used as an initial screening step before applying a secondary technique of greater discriminatory power [42, 47, 48].

Mycobacterial interspersed repetitive unit variable number of tandem repeats and other polymerase chain reaction-based techniques

A variety of other PCR-based techniques for typing *M. tuberculosis* isolates have been developed. In a recent interlaboratory comparative study, six such techniques were compared for their discriminatory power and reproducibility [46]. These techniques were easy and quick to perform but only two of them, mixed-linker PCR and the variable number of tandem repeats (VNTR) method were sufficiently reproducible [44, 49]. However, of these two techniques, only mixed-linker PCR was comparable to IS*6110* typing in terms of discriminatory power. Unfortunately, mixed-linker PCR is also a technique that depends on IS*6110*-generated polymorphisms, and so is of limited use amongst low-band-number isolates. This is also the case with another promising IS*6110*-based PCR technique, ligation-mediated PCR [46].

A more promising approach for developing a PCR-based typing system is to identify novel polymorphic loci, which are independent of existing techniques such as IS*6110* typing. Mycobacterial interspersed repetitive units (MIRUs) are an example of such elements [50–52]. They are a specific class of VNTR that have been identified at 41 different loci in the genome of *M. tuberculosis*. Each comprises strings of short repetitive sequences (<100 base pairs). The number of repeats at different loci varies between strains. PCR amplification across each MIRU, therefore, generates fragments of different sizes from different strains. If these fragments are accurately sized, the number of repeats at each loci can be determined. Analysing 12 of the most hypervariable loci resulted in a discriminative power close to that of IS*6110* typing. This type of approach is particularly suited for use with global databases as each typed strain is assigned a 12-digit number corresponding to the number of repeats at each MIRU locus [53]. This unambiguous coding system makes interlaboratory comparisons facile. The main limitation of MIRU typing is the technical difficulty associated with accurately sizing multiple small PCR fragments. This can be partly overcome by combining multiplex PCR with a fluorescence-based DNA analyser [52].

Molecular markers and phylogenetics

Most attempts at developing typing techniques for the study of tuberculosis have been aimed at distinguishing between individual strains of *M. tuberculosis* in order to define chains of transmission. The genetic mechanisms underlying such techniques need to be highly polymorphic, with rapid molecular clocks, to generate sufficient diversity amongst strains. However, these techniques are not particularly well suited to studying phylogenetic relationships between more distantly related strains. For example IS*6110* copy number showed no relationship to a phylogeny of diverse *M. tuberculosis* strains based on single nucleotide polymorphisms in housekeeping genes [54]. Interest in defining these phylogenetic relationships has developed recently with the identification of groups of closely related strains, such as the Beijing family (see below), that appear to have a specific phenotype.

Deletion analysis is a particularly attractive approach to studying the phylogeny of *M. tuberculosis* strains as it could simultaneously provide information about the biological basis of a unique strain phenotype [55]. In essence, the genome of a strain is evaluated using a microarray to determine whether or not any deletions have occurred relative to the sequenced reference strain. Since these deletions rarely occur independently at exactly the same chromosomal locus, they can be seen as unique and irreversible genetic events. The number and distribution of these deletions, therefore, provides a genomic signature which can be used for constructing robust phylogenetic relationships. These deletions disrupt coding regions of the genome and it is tempting to speculate that the loss of specific genes could influence important characteristics of strains such as transmissibility or antigenicity [3, 56].

Lessons from molecular epidemiology

Dynamics of transmission within populations

Molecular epidemiological investigations have been very useful in providing an understanding of the

transmission dynamics of tuberculosis within a community. These studies are based upon the premise that patients infected with strains showing identical fingerprints, termed "clustered cases", are the result of recent transmission, whereas those infected with isolates with unique RFLP patterns are presumed to represent remote transmission and thus reactivation of strains acquired in the more distant past.

Before the availability of more reliable molecular epidemiological tools, the belief was that 10% of patients developed *M. tuberculosis* disease as a result of recent transmission. However, population-based molecular epidemiological studies in San Francisco, (CA, USA), New York (NY, USA) and Amsterdam (the Netherlands) have refined the understanding of this subject. These studies demonstrated that the rate of recent infection was much higher than the estimated 10% predicted by traditional epidemiological studies. For example, in San Francisco, almost one-third of new cases of tuberculosis were as a result of recent infection. In New York, clustering was estimated to be 40% and, in Amsterdam, 47% [57–59]. In San Francisco, one patient was found to be the index case and accounted for the transmission of 6% of new cases of tuberculosis in the city during 1991–1992. In these studies, the risk factors associated with recent transmission were lower socioeconomic group, native ethnic minority and acquired immune deficiency syndrome. These studies have important ramifications for tuberculosis control. They demonstrate that ongoing transmission of infection contributed to the disease burden at much higher rates than previously thought, and highlighted the importance of control efforts in interrupting transmission, especially amongst groups at high risk.

Population-based studies in Norway and Switzerland showed percentages of clustering that were relatively low (16 and 17.5% respectively) compared to the other studies [60, 61]. This low level of recent transmission suggested that tuberculosis control was more effective in these settings. RFLP fingerprinting studies, therefore, can be used as a tool to monitor the performance of a tuberculosis control programme. In addition, this type of study can be used to identify specific risk factors for tuberculosis transmission and assist in targeting interventions to the subpopulations that disproportionately contribute to transmission. For example, in San Francisco, intensification of tuberculosis control decreased the overall numbers of those with recent infection [62] but also demonstrated persistent transmission among difficult-to-target high-risk groups.

The general principle that clustering of IS*6110* patterns equates with recent transmission might not always apply in all situations. For instance, studies have shown that the rate of clustering varies depending on the area under study. In resource-poor countries, clustering rates of 14–41% were found [9, 29, 63–66]. These findings were unexpected since the rates of clustering in some of these studies are comparable to or lower than those seen in low incidence countries. In these higher-incidence areas, with poor tuberculosis control, more transmission is expected. However, most of these studies do not report the incidence of tuberculosis, and, in many cases, it is difficult to ascertain the percentage of samples analysed in the study of those available in the community. This is particularly important because the number of isolates sampled in the study may be small in comparison to the total number of circulating isolates contributing to transmission in the area. Since the degree of sampling affects clustering, the rates of transmission may have been grossly underestimated [67, 68]. Consequently, it is imperative, in community studies, to include a high percentage of circulating isolates in order to accurately determine the rates of recent and reactivated disease.

Epidemiological impact of subpopulations on tuberculosis transmission

The impact of immigrant subpopulations on the epidemiology of tuberculosis in the population of a community has been perceived as an important public health issue in developed countries. For example, during 1986–1997, the number of tuberculosis cases diagnosed in foreign-born persons in the USA increased by 56%. These statistics suggest that immigrants could be transmitting tuberculosis to the native population. However, in one study, only one of 43 cases amongst immigrants resulted in two secondary cases of tuberculosis infection in US-born cases. Additionally one-fifth of Mexican-born patients acquired their tuberculosis infection in San Francisco [69]. This study was particularly important because the native population were proved to have transmitted tuberculosis at a higher rate to immigrants than immigrants to natives. A study from the same city described two parallel epidemiological patterns of tuberculosis in foreign-born and US-born populations [70]. Most foreign-born individuals develop tuberculosis from reactivation, whereas 20% of US-born cases developed tuberculosis from recent infection. In other settings, different patterns of transmission have been described within immigrant populations. Tuberculosis transmission has been documented from immigrants in Denmark [71, 72]. In the Netherlands, almost half of the recent transmission occurs in immigrant groups, but most of the transmission occurs within the same nationality [73].

The transmission index, defined as the mean number of tuberculosis cases resulting from recent transmission of a potential source case, has been used to quantify transmission between different subpopulations [73]. In San Francisco, the transmission index was found to be lower among foreign-born than US-born patients and was much higher among Black American patients of <35 yrs [74]. These studies show that defining the nature of transmission between different population groups can be used to inform and strengthen tuberculosis control.

Epidemiological suspected and unsuspected transmission

In addition to the studies described above, molecular epidemiology has contributed to improving

disease control in other ways. For example one study demonstrated the explosive potential for tuberculosis to progress to disease and spread amongst HIV-infected persons [75]. In this study, conventional surveillance detected 12 cases of tuberculosis in a residential facility for persons with HIV disease. Analysis of isolates by IS*6110* RFLP demonstrated that newly acquired tuberculous infection in HIV-infected patients spread readily and progressed within 3 months of exposure to disease, demonstrating the particular vulnerability of HIV-infected individuals to exogenous tuberculosis infection.

Molecular epidemiology has also documented the potential for spread of drug-resistant strains among hospitalised patients [14, 76–79]. During one 43-month period, New York City accounted for almost one-quarter of all cases of MDR-TB in the USA. Most of these patients were infected with HIV and were found to have acquired their often-fatal MDR-TB whilst in hospital [80]. The results of this and other similar studies have led to more rigorous adherence to infection control policies, particularly in settings in which there are many HIV-infected persons [81]. Increased surveillance with prompt diagnosis and appropriate therapy in settings such as hospitals, prisons, schools and homeless shelters is now resulting in an overall decrease in tuberculosis transmission [81].

Population-based molecular typing studies have also shown the dramatic impact that an individual or a small group of individuals can have on tuberculosis transmission. In Minneapolis (MN, USA), during 1992, one individual was shown to have caused 35% of all new active cases of tuberculosis [80]. One of the first large-scale molecular epidemiological studies uncovered extensive transmission of *M. tuberculosis* among a small group of substance abusers with significant "spillover" to the general population [82]. Knowledge of the negative impact that poorly managed patients can have in a community emphasises the need for a sustained level of tuberculosis control for each and every disease case.

Active case finding, through the evaluation of individuals who have been in contact with infectious tuberculosis patients, is a traditional activity of tuberculosis control programmes in industrialised countries. A basic principle of this policy is that contacts are likely to have been infected by the infectious case and thus carry the same strain (with the same drug susceptibility pattern) as the index case. This hypothesis was tested in a study of index and contact cases [83]. The authors compared the DNA fingerprints of pairs of indexes and contacts who were both ultimately diagnosed with tuberculosis. Thirty per cent (16 of 54) of pairs had different fingerprints, demonstrating that the contact had been infected from an unidentified third person. This illustrates that transmission links are often more complex than those assumed by conventional epidemiology. Similarly, a contact investigation among five large clusters in the Netherlands showed transmission occurred after only transient contact, contrary to the conventional view that tuberculosis is usually acquired following prolonged exposure to an infectious case [84]. This complexity is also seen in population-based molecular studies. For example it is often difficult to establish an epidemiological link between individual cases in an RFLP cluster. Fingerprinting and contact investigation in different settings demonstrated epidemiological links in only 5–10% of cases [58–60, 84].

Identical fingerprints do not always correlate with an established epidemiological link. For example strains from a rural area in which patients were geographically dispersed and highly unlikely to have had previous contact were found to have identical fingerprints [33]. Also, strains from cases in different states in the USA were found to be identical, although they had no history of previously having been in contact [26]. Similarly, in South Africa, Tanzania, Zimbabwe, Kenya and Malawi, strains from different dispersed geographical regions were found to have identical patterns [85]. These studies show that some strains are more prevalent and abundant than others and this could lead to the misattribution of epidemiological links. Therefore, caution should be exercised in reaching conclusions based solely on identical fingerprint results.

The interpretation of molecular epidemiological studies should largely depend on the study question, the area under study and the typing methods used. A combination of typing methods based on more rapid and slower molecular clocks should, in principle, be able to differentiate between the contributions of remote and recent transmission to clustering. Typing methods with a reliably slow molecular clock could be very useful in global tracking and evolutionary studies of tuberculosis. Thus, the future challenge is to identify techniques with different molecular clocks that can identify, with certainty, recent and remote evolutionary linkage between strains.

Quantification of the level of infectiousness among smear-negative patients

It is generally believed that patients with tuberculosis whose sputum microscopic examination fails to detect acid-fast bacilli (AFB) are significantly less infectious than those with positive smears. However, a molecular epidemiological study that compared transmission from AFB smear-positive and -negative patients suggested that AFB smear-negative patients were responsible for $\geqslant 21\%$ of tuberculosis transmitted in the city of San Francisco [86]. Thus intensifying tuberculosis control measures for smear-negative cases could significantly reduce the transmission of tuberculosis.

False-positive Mycobacterium tuberculosis *cultures*

There are now a plethora of studies describing the problem of false-positive *M. tuberculosis* cultures in the laboratory with the use of multiple markers [87–94]. Laboratory cross-contamination represents a significant problem for the microbiologist and may result in unnecessary treatment and potential

drug toxicity for a patient. For instance, a small but significant proportion (3%) of New York City patients had falsely positive cultures for *M. tuberculosis* as a result of contamination [79]. Timely molecular analysis and appropriate changes to specimen processing have been identified as useful measures for avoiding false-positive cultures [95, 96].

Current topics of interest

Epidemiological prevalence of the Beijing genotype

A distinct family of *M. tuberculosis* strains (subsequently labelled the "W strain") was associated with >350 cases in New York City and, at one point, accounted for 25% of all MDR-TB cases in the USA [97]. These strains were later demonstrated to belong to a branch of a distinct family of strains named the "Beijing genotype" because of their predominance in the Beijing area of China. Although the IS*6110* RFLP pattern of the Beijing strains is not unique and is often difficult to distinguish from other genotypes, the spoligotype is highly distinctive and easy to identify. With the use of spoligotyping, the Beijing strains have been associated with transmission in Azerbaijan, Thailand, Estonia, Iran, Vietnam, Malaysia, Estonia, China, Hong Kong, South Africa, Colombia and Russia [48, 98–103]. Its presence has been detected throughout Southeast Asia and in Hong Kong, and, in one area of study, the Beijing genotype family accounted for 70% of all isolates [104].

On the island of Gran Canaria, this genotype was initially introduced by an immigrant from Africa. Over a period of 4 yrs after its introduction, this genotype became the most common isolate on the island [105]. In some areas of the world, this strain has been associated with cases of tuberculosis in young individuals [48, 106]. The association of this genotype with a younger age group is recognised as an indication of ongoing transmission. In addition, studies have found a significant correlation between the Beijing genotype and drug resistance [97, 103, 107]. The global dissemination and apparent transmissibility of this strain has raised the tantalising possibility that these epidemiological characteristics are a reflection of an intrinsic biological property unique to this family. It has also been postulated that, if these strains can indeed replicate more efficiently in the host, this could favour the development of resistance [105]. Intensive research efforts are now being directed towards elucidating the genetic basis of this apparent phenotype. However, it remains to be seen whether or not this family of strains is more virulent and what its role may be in the worldwide tuberculosis epidemic.

Epidemiological evidence of exogenous reinfection

It was not until molecular fingerprinting techniques became available that the exogenous reinfection phenomenon was conclusively demonstrated to occur. An evaluation of HIV-infected patients in a New York City hospital, who repeatedly yielded positive cultures for *M. tuberculosis*, identified 11 patients with sequential isolates that became resistant to antimicrobial agents. In four of these patients, the RFLP patterns of the isolates changed dramatically at the time that drug resistance was detected. In these patients, the clinical and microbiological evidence was consistent with the presence of active tuberculosis caused by a new strain of *M. tuberculosis* [107]. Exogenous reinfection of persons whose only immunosuppressing condition was diabetes has also been demonstrated [108].

A small study in Africa demonstrated the importance of considering reinfection in patients with a past history of tuberculosis. Original and recurrent isolates of tuberculosis were analysed in five patients in this study who were known to be HIV-positive with recurrence of tuberculosis [109]. Reinfection was demonstrated in one patient whose original and recurrent isolates had dramatically different fingerprints. This study demonstrates that, when tuberculosis recurs after standard treatment in HIV patients, reinfection with a new isolate and not just relapse should be considered.

In high-incidence settings, particularly those with high rates of HIV coinfection, it is anticipated that re-infection is important. This was the case in a study in South Africa in which up to 75% (12 of 16) of patients had different fingerprints in their initial and second episodes of disease [110]. However, in another study carried out in South African miners, only 2% (1 of 48) of cases were found to be due to reinfection [109]. Other studies in Hong Kong [8] and India [111] reported rates of reinfection ranging 12–31%. In these studies, different methodologies and possible contamination could have accounted for the discrepancies in the results. Thus, it is difficult to draw definite conclusions as to the rate of reinfection in high-incidence countries based on these studies.

In a study on the island of Gran Canaria, Spain, in a setting with good tuberculosis control measures and in conjunction with meticulous methodology, 44% (8 of 18) of cases were attributed to reinfection. Thus, even in communities with a low incidence of tuberculosis, reinfection appears to be important [112].

Accurately establishing the rates of reinfection in different settings is important in predicting the effects of control strategies, such as directly observed treatment, short course (DOTS), on the course of the current tuberculosis epidemic. For example, if DOTS rapidly shuts off transmission and reinfection is rare, the epidemic will die off slowly because cases will continue to appear through reactivation. Conversely, if reactivation is not common, the epidemic will decline relatively quickly as DOTS prevents new cases by stopping reinfection. Further, determining the rates of reinfection *versus* reactivation in areas with high levels of HIV coinfection is of critical importance for designing the most appropriate chemoprophylactic strategies. For example, in an area with high levels of exogenous reinfection, short-term chemoprophylaxis is unlikely to be of value.

The relative fitness of isoniazid-resistant strains

The development and transmission of MDR-TB, defined as resistance to at least isoniazid and rifampicin, is of particular concern because it requires prolonged costly therapy that often produces only low cure rates [113, 114]. In spite of the implications of the spread of drug-resistant tuberculosis, little is known about the ability of drug-resistant strains to transmit, survive and reproduce in a population as compared to susceptible strains [115].

Experimental models suggest that antibiotic resistance imposes a biological cost on bacterial fitness [116, 117]. It is thought that the mechanism of resistance, for example to isoniazid (mutations in the *katG* gene), causes an intrinsic cost to the fitness of the microorganism to be incurred [115]. This is supported by *in vitro* studies that show that *katG* is required for optimal survival of *M. tuberculosis* in animal models [118].

Additional evidence that resistant strains are less fit comes from molecular epidemiological studies carried out in the Netherlands, Mexico and South Africa, in which less clustering was observed among drug-resistant strains [84, 109, 119]. In a more in-depth analysis of clustering in San Francisco, the case reproduction number of circulating *Mycobacterium tuberculosis* strains resistant to isoniazid was found to be significantly reduced when compared to the case reproduction number of susceptible circulating phenotypes (data not shown). These molecular epidemiological studies suggest that development of isoniazid resistance incurs a significant biological cost for strains of *M. tuberculosis*. This decreased fitness could diminish the epidemiological impact of isoniazid-resistant tuberculosis and MDR-TB phenotypes in a population. However, it remains to be seen whether the magnitude of the fitness cost can overcome the many factors that collude to promote the transmission of drug-resistant tuberculosis in communities with a high burden of tuberculosis.

Conclusion

The development of molecular tools has added a new dimension to the classical epidemiology of tuberculosis and greatly enhanced understanding of the complex transmission dynamics within populations and between hosts. In the process, molecular epidemiology has demonstrated inadequacies in tuberculosis control programmes and helped accumulate motivation and resources for their improvement. Other technologies based on knowledge of the complete genome sequence of *M. tuberculosis*, which will provide newer tools for probing the epidemiology of tuberculosis, are now emerging. In spite of recent research advances, tuberculosis continues to remain a devastating infectious disease, disproportionately impacting on the world′s poorest countries. The future challenge for molecular epidemiology is to provide better understanding of the transmisson dynamics of tuberculosis in those countries with the greatest burden of disease, and to stimulate an urgency to improving control measures on a more global scale.

References

1. World Health Organization. Global Tuberculosis Control. WHO Report 2001. Geneva, World Health Organization, 2001.
2. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *JAMA* 1999; 282: 677–686.
3. Kato-Maeda M, Rhee JT, Gingeras TR, *et al.* Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res* 2001; 11: 547–554.
4. Cole ST, Brosch R, Parkhill J, *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998; 393: 537–544.
5. Thierry D, Brisson-Noel A, Vincent-Levy-Frebault V, Nguyen S, Guesdon JL, Gicquel B. Characterization of a *Mycobacterium tuberculosis* insertion sequence, IS*6110*, and its application in diagnosis. *J Clin Microbiol* 1990; 28: 2668–2673.
6. Small PM, Shafer RW, Hopewell PC, *et al.* Exogenous reinfection with multidrug-resistant *Mycobacterium tuberculosis* in patients with advanced HIV infection. *N Engl J Med* 1993; 328: 1137–1144.
7. van Soolingen D, Hermans PW, de Haas PE, Soll DR, van Embden JD. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol* 1991; 29: 2578–2586.
8. Das S, Chan SL, Allen BW, Mitchison DA, Lowrie DB. Application of DNA fingerprinting with IS*986* to sequential mycobacterial isolates obtained from pulmonary tuberculosis patients in Hong Kong before, during and after short-course chemotherapy. *Tuber Lung Dis* 1993; 74: 47–51.
9. Yuen LK, Ross BC, Jackson KM, Dwyer B. Characterization of *Mycobacterium tuberculosis* strains from Vietnamese patients by Southern blot hybridization. *J Clin Microbiol* 1993; 31: 1615–1618.
10. van Soolingen D, de Haas PE, Hermans PW, Groenen PM, van Embden JD. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. *J Clin Microbiol* 1993; 31: 1987–1995.
11. Agasino CB, Ponce de Leon A, Jasmer RM, Small PM. Epidemiology of *Mycobacterium tuberculosis* strains in San Francisco that do not contain IS*6110*. *Int J Tuberc Lung Dis* 1998; 2: 518–520.
12. Hermans PW, van Soolingen D, Dale JW, *et al.* Insertion element IS*986* from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. *J Clin Microbiol* 1990; 28: 2051–2058.
13. van Embden JD, van Soolingen D, Small PM, Hermans PW. Genetic markers for the epidemiology of tuberculosis. *Res Microbiol* 1992; 143: 385–391.
14. Ritacco V, Di Lonardo M, Reniero A, *et al.* Nosocomial spread of human immunodeficiency virus-related multidrug-resistant tuberculosis in Buenos Aires. *J Infect Dis* 1997; 176: 637–642.
15. Samper S, Martin C, Pinedo A, *et al.* Transmission between HIV-infected patients of multidrug-resistant

tuberculosis caused by *Mycobacterium bovis*. *AIDS* 1997; 11: 1237–1242.
16. Namwat W, Luangsuk P, Palittapongarnpim P. The genetic diversity of *Mycobacterium tuberculosis* strains in Thailand studied by amplification of DNA segments containing direct repetitive sequences. *Int J Tuberc Lung Dis* 1998; 2: 153–159.
17. Gillespie SH, Kennedy N, Ngowi FI, Fomukong NG, al-Maamary S, Dale JW. Restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolated from patients with pulmonary tuberculosis in Northern Tanzania. *Trans R Soc Trop Med Hyg* 1995; 89: 335–338.
18. Rasolofo-Razanamparany V, Ramarokoto H, Auregan G, Gicquel B, Chanteau S. A combination of two genetic markers is sufficient for restriction fragment length polymorphism typing of *Mycobacterium tuberculosis* complex in areas with a high incidence of tuberculosis. *J Clin Microbiol* 2001; 39: 1530–1535.
19. Sola C, Filliol I, Legrand E, Mokrousov I, Rastogi N. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS*1081*, IS*6110*, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. *J Mol Evol* 2001; 53: 680–689.
20. Barlow RE, Gascoyne-Binzi DM, Gillespie SH, Dickens A, Qamer S, Hawkey PM. Comparison of variable number tandem repeat and IS*6110*-restriction fragment length polymorphism analyses for discrimination of high- and low-copy-number IS*6110 Mycobacterium tuberculosis* isolates. *J Clin Microbiol* 2001; 39: 2453–2457.
21. Yang Z, Chaves F, Barnes PF, *et al.* Evaluation of method for secondary DNA typing of *Mycobacterium tuberculosis* with pTBN12 in epidemiologic study of tuberculosis. *J Clin Microbiol* 1996; 34: 3044–3048.
22. Singh SP, Salamon H, Lahti CJ, Farid-Moyer M, Small PM. Use of pulsed-field gel electrophoresis for molecular epidemiologic and population genetic studies of *Mycobacterium tuberculosis*. *J Clin Microbiol* 1999; 37: 1927–1931.
23. Yeh RW, Ponce de Leon A, Agasino CB, *et al.* Stability of *Mycobacterium tuberculosis* DNA genotypes. *J Infect Dis* 1998; 177: 1107–1111.
24. de Boer AS, Borgdorff MW, de Haas PE, Nagelkerke NJ, van Embden JD, van Soolingen D. Analysis of rate of change of IS*6110* RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. *J Infect Dis* 1999; 180: 1238–1244.
25. Behr MA, Wilson MA, Gill WP, *et al.* Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 1999; 284: 1520–1523.
26. Yang ZH, Bates JH, Eisenach KD, Cave MD. Secondary typing of *Mycobacterium tuberculosis* isolates with matching IS*6110* fingerprints from different geographic regions of the United States. *J Clin Microbiol* 2001; 39: 1691–1695.
27. Burman WJ, Reves RR, Hawkes AP, *et al.* DNA fingerprinting with two probes decreases clustering of *Mycobacterium tuberculosis*. *Am J Respir Crit Care Med* 1997; 155: 1140–1146.
28. Warren R, Richardson M, Sampson S, *et al.* Genotyping of *Mycobacterium tuberculosis* with additional markers enhances accuracy in epidemiological studies. *J Clin Microbiol* 1996; 34: 2219–2224.
29. Hermans PW, Messadi F, Guebrexabher H, *et al.* Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and the Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. *J Infect Dis* 1995; 171: 1504–1513.
30. Hermans PW, van Soolingen D, van Embden JD. Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*. *J Bacteriol* 1992; 174: 4157–4165.
31. Poulet S, Cole ST. Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch Microbiol* 1995; 163: 87–95.
32. Poulet S, Cole ST. Repeated DNA sequences in mycobacteria. *Arch Microbiol* 1995; 163: 79–86.
33. Braden CR, Templeton GL, Cave MD, *et al.* Interpretation of restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates from a state with a large rural population. *J Infect Dis* 1997; 175: 1446–1452.
34. Xet-Mull AM, Kubin M, Prikazsky V, Harris EH, Riley LW. DRE-PCR (double repetitive element-polymerase chain reaction) - a new molecular-epidemiologic method in the detection of tuberculosis. *Epidemiol Mikrobiol Imunol* 2001; 50: 82–86.
35. Suffys P, Vanderborght PR, Santos PB, Correa LA, Bravin Y, Kritski AL. Inhibition of the polymerase chain reaction by sputum samples from tuberculosis patients after processing using a silica-guanidiniumthiocyanate DNA isolation procedure. *Mem Inst Oswaldo Cruz* 2001; 96: 1137–1139.
36. Arias-Bouda LP, Kolk AH. PCR-based assays for the diagnosis of tuberculosis. *Int J Tuberc Lung Dis* 2001; 5: 1163–1164.
37. Bogard M, Vincelette J, Antinozzi R, *et al.* Multicenter study of a commercial, automated polymerase chain reaction system for the rapid detection of *Mycobacterium tuberculosis* in respiratory specimens in routine clinical practice. *Eur J Clin Microbiol Infect Dis* 2001; 20: 724–731.
38. Tansuphasiri U. Detection of *Mycobacterium tuberculosis* from sputum collected on filter paper and stored at room temperature for 5 days by PCR assay and culture. *J Med Assoc Thai* 2001; 84: 1183–1189.
39. Kamerbeek J, Schouls L, Kolk A, *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 1997; 35: 907–914.
40. Goyal M, Lawn S, Afful B, Acheampong JW, Griffin G, Shaw R. Spoligotyping in molecular epidemiology of tuberculosis in Ghana. *J Infect* 1999; 38: 171–175.
41. Samper S, Otal I, Rubio MC, Vitoria MA, Gomez-Lus R, Martin C. Application of RFLP to typing strains of *Mycobacterium tuberculosis*. *Enferm Infecc Microbiol Clin* 1993; 11: 547–551.
42. Goguet de la Salmoniere YO, Li HM, Torrea G, Bunschoten A, van Embden J, Gicquel B. Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*. *J Clin Microbiol* 1997; 35: 2210–2214.
43. Friedman CR, Stoeckle MY, Johnson WD Jr, Riley LW. Double-repetitive-element PCR method for subtyping *Mycobacterium tuberculosis* clinical isolates. *J Clin Microbiol* 1995; 33: 1383–1384.
44. Haas WH, Butler WR, Woodley CL, Crawford JT. Mixed-linker polymerase chain reaction: a new method for rapid fingerprinting of isolates of the

Mycobacterium tuberculosis complex. *J Clin Microbiol* 1993; 31: 1293–1298.
45. Molhuizen HO, Bunschoten AE, Schouls LM, van Embden JD. Rapid detection and simultaneous strain differentiation of *Mycobacterium tuberculosis* complex bacteria by spoligotyping. *Methods Mol Biol* 1998; 101: 381–394.
46. Kremer K, van Soolingen D, Frothingham R, *et al.* Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J Clin Microbiol* 1999; 37: 2607–2618.
47. Diaz R, Kremer K, de Haas PE, *et al.* Molecular epidemiology of tuberculosis in Cuba outside of Havana, July 1994–June 1995: utility of spoligotyping *versus* IS*6110* restriction fragment length polymorphism. *Int J Tuberc Lung Dis* 1998; 2: 743–750.
48. Doroudchi M, Kremer K, Basiri EA, Kadivar MR, Van Soolingen D, Ghaderi AA. IS*6110*-RFLP and spoligotyping of *Mycobacterium tuberculosis* isolates in Iran. *Scand J Infect Dis* 2000; 32: 663–668.
49. Skuce RA, McCorry TP, McCarroll JF, *et al.* Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. *Microbiology* 2002; 148: 519–528.
50. Supply P, Magdalena J, Himpens S, Locht C. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Mol Microbiol* 1997; 26: 991–1003.
51. Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Locht C. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol* 2000; 36: 762–771.
52. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol* 2001; 39: 3563–3571.
53. Mazars E, Lesjean S, Banuls AL, *et al.* High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci USA* 2001; 98: 1901–1906.
54. Sreevatsan S, Pan X, Stockbauer KE, *et al.* Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci USA* 1997; 94: 9869–9874.
55. Fitzgerald JR, Sturdevant DE, Mackie SM, Gill SR, Musser JM. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc Natl Acad Sci USA* 2001; 98: 8821–8826.
56. Salamon H, Kato-Maeda M, Small PM, Drenkow J, Gingeras TR. Detection of deleted genomic DNA using a semiautomated computational analysis of GeneChip data. *Genome Res* 2000; 10: 2044–2054.
57. Alland D, Kalkut GE, Moss AR, *et al.* Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N Engl J Med* 1994; 330: 1710–1716.
58. Small PM, Hopewell PC, Singh SP, *et al.* The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med* 1994; 330: 1703–1709.
59. van Deutekom H, Gerritsen JJ, van Soolingen D, van Ameijden EJ, van Embden JD, Coutinho RA. A molecular epidemiological approach to studying the transmission of tuberculosis in Amsterdam. *Clin Infect Dis* 1997; 25: 1071–1077.
60. Pfyffer GE, Strassle A, Rose N, Wirth R, Brandli O, Shang H. Transmission of tuberculosis in the metropolitan area of Zurich: a 3 year survey based on DNA fingerprinting. *Eur Respir J* 1998; 11: 804–808.
61. Heldal E, Docker H, Caugant DA, Tverdal A. Pulmonary tuberculosis in Norwegian patients. The role of reactivation, re-infection and primary infection assessed by previous mass screening data and restriction fragment length polymorphism analysis. *Int J Tuberc Lung Dis* 2000; 4: 300–307.
62. Chin DP, Crane CM, Diul MY, *et al.* Spread of *Mycobacterium tuberculosis* in a community implementing recommended elements of tuberculosis control. *JAMA* 2000; 283: 2968–2974.
63. Godfrey-Faussett P, Mortimer PR, Jenkins PA, Stoker NG. Evidence of transmission of tuberculosis by DNA fingerprinting. *BMJ* 1992; 305: 221–223.
64. Chevrel-Dellagi D, Abderrahman A, Haltiti R, Koubaji H, Gicquel B, Dellagi K. Large-scale DNA fingerprinting of *Mycobacterium tuberculosis* strains as a tool for epidemiological studies of tuberculosis. *J Clin Microbiol* 1993; 31: 2446–2450.
65. Gomez Marin JE, Rigouts L, Villegas Londono LE, Portaels F. Restriction fragment length polymorphism (RFLP) analysis and tuberculosis epidemiology. *Bull Pan Am Health Organ* 1995; 29: 226–236.
66. Dale JW, Nor RM, Ramayah S, Tang TH, Zainuddin ZF. Molecular epidemiology of tuberculosis in Malaysia. *J Clin Microbiol* 1999; 37: 1265–1268.
67. Glynn JR, Vynnycky E, Fine PE. Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. *Am J Epidemiol* 1999; 149: 366–371.
68. Murray M. Determinants of cluster distribution in the molecular epidemiology of tuberculosis. *Proc Natl Acad Sci USA* 2002; 99: 1538–1543.
69. Jasmer RM, Ponce de Leon A, Hopewell PC, *et al.* Tuberculosis in Mexican-born persons in San Francisco: reactivation, acquired infection and transmission. *Int J Tuberc Lung Dis* 1997; 1: 536–541.
70. Chin DP, DeRiemer K, Small PM, *et al.* Differences in contributing factors to tuberculosis incidence in US-born and foreign-born persons. *Am J Respir Crit Care Med* 1998; 158: 1797–1803.
71. Yang ZH, de Haas PE, van Soolingen D, van Embden JD, Andersen AB. Restriction fragment length polymorphism *Mycobacterium tuberculosis* strains isolated from Greenland during 1992: evidence of tuberculosis transmission between Greenland and Denmark. *J Clin Microbiol* 1994; 32: 3018–3025.
72. Yang ZH, de Haas PE, Wachmann CH, van Soolingen D, van Embden JD, Andersen AB. Molecular epidemiology of tuberculosis in Denmark in 1992. *J Clin Microbiol* 1995; 33: 2077–2081.
73. Borgdorff MW, Nagelkerke N, van Soolingen D, de Haas PE, Veen J, van Embden JD. Analysis of tuberculosis transmission between nationalities in the Netherlands in the period 1993–1995 using DNA fingerprinting. *Am J Epidemiol* 1998; 147: 187–195.
74. Borgdorff MW, Behr MA, Nagelkerke NJ, Hopewell PC, Small PM. Transmission of tuberculosis in San

Francisco and its association with immigration and ethnicity. *Int J Tuberc Lung Dis* 2000; 4: 287–294.
75. Daley CL, Small PM, Schecter GF, *et al.* An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus. An analysis using restriction-fragment-length polymorphisms. *N Engl J Med* 1992; 326: 231–235.
76. Anastasis D, Pillai G, Rambiritch V, Abdool Karim SS. A retrospective study of human immunodeficiency virus infection and drug-resistant tuberculosis in Durban, South Africa. *Int J Tuberc Lung Dis* 1997; 1: 220–224.
77. Angarano G, Carbonara S, Costa D, Gori A. Drug-resistant tuberculosis in human immunodeficiency virus infected persons in Italy. *Int J Tuberc Lung Dis* 1998; 2: 303–311.
78. Breathnach AS, de Ruiter A, Holdsworth GM, *et al.* An outbreak of multi-drug-resistant tuberculosis in a London teaching hospital. *J Hosp Infect* 1998; 39: 111–117.
79. Frieden TR, Woodley CL, Crawford JT, Lew D, Dooley SM. The molecular epidemiology of tuberculosis in New York City: the importance of nosocomial transmission and laboratory error. *Tuber Lung Dis* 1996; 77: 407–413.
80. Frieden TR, Sherman LF, Maw KL, *et al.* A multi-institutional outbreak of highly drug-resistant tuberculosis: epidemiology and clinical outcomes. *JAMA* 1996; 276: 1229–1235.
81. Fujiwara PI, Cook SV, Rutherford CM, *et al.* A continuing survey of drug-resistant tuberculosis, New York City, April 1994. *Arch Intern Med* 1997; 157: 531–536.
82. Genewein A, Telenti A, Bernasconi C, *et al.* Molecular approach to identifying route of transmission of tuberculosis in the community. *Lancet* 1993; 342: 841–844.
83. Behr MA, Hopewell PC, Paz EA, Kawamura LM, Schecter GF, Small PM. Predictive value of contact investigation for identifying recent transmission of *Mycobacterium tuberculosis. Am J Respir Crit Care Med* 1998; 158: 465–469.
84. van Soolingen D, Borgdorff MW, de Haas PE *et al.* Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis* 1999; 180: 726–736.
85. Picard B. Molecular epidemiology of large bacterial endemics in sub-Saharan Africa. *Bull Soc Pathol Exot* 2000; 93: 219–223.
86. Behr MA, Warren SA, Salamon H, *et al.* Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *Lancet* 1999; 353: 444–449.
87. Chang CL, Kim HH, Son HC, *et al.* False-positive growth of *Mycobacterium tuberculosis* attributable to laboratory contamination confirmed by restriction fragment length polymorphism analysis. *Int J Tuberc Lung Dis* 2001; 5: 861–867.
88. Breese PE, Burman WJ, Hildred M, *et al.* The effect of changes in laboratory practices on the rate of false-positive cultures for *Mycobacterium tuberculosis. Arch Pathol Lab Med* 2001; 125: 1213–1216.
89. Gascoyne-Binzi DM, Barlow RE, Frothingham R, *et al.* Rapid identification of laboratory contamination with *Mycobacterium tuberculosis* using variable number tandem repeat analysis. *J Clin Microbiol* 2001; 39: 69–74.
90. Burman WJ, Reves RR. Review of false-positive cultures for *Mycobacterium tuberculosis* and recommendations for avoiding unnecessary treatment. *Clin Infect Dis* 2000; 31: 1390–1395.
91. Nivin B, Driscoll J, Glaser T, Bifani P, Munsiff S. Use of spoligotype analysis to detect laboratory cross-contamination. *Infect Control Hosp Epidemiol* 2000; 21: 525–527.
92. Anon. Misdiagnoses of tuberculosis resulting from laboratory cross-contamination of *Mycobacterium tuberculosis* cultures - New Jersey, 1998. *MMWR Morb Mortal Wkly Rep* 2000; 49: 413–416.
93. Perfecto B, Dorronsoro I, Lopez-Goni I. Confirmation by molecular typing of cross-contamination in a mycobacteria laboratory. *Enferm Infecc Microbiol Clin* 2000; 18: 12–15.
94. Carricajo A, Vincent V, Berthelot P, Gery P, Aubert G. Mycobacterial cross-contamination of bronchoscope detected by molecular techniques. *J Hosp Infect* 1999; 42: 252–253.
95. Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, Mickelsen PA. Molecular strain typing of *Mycobacterium tuberculosis* to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. *J Clin Microbiol* 1993; 31: 1677–1682.
96. Behr MA, Small PM. Molecular fingerprinting of *Mycobacterium tuberculosis*: how can it help the clinician? *Clin Infect Dis* 1997; 25: 806–810.
97. Bifani PJ, Mathema B, Liu Z, *et al.* Identification of a W variant outbreak of *Mycobacterium tuberculosis via* population-based molecular epidemiology. *JAMA* 1999; 282: 2321–2327.
98. Pfyffer GE, Strassle A, van Gorkum T, *et al.* Multidrug-resistant tuberculosis in prison inmates, Azerbaijan. *Emerg Infect Dis* 2001; 7: 855–861.
99. van Soolingen D, Kremer K, Borgdorff M. *Mycobacterium tuberculosis* Beijing genotype, Thailand – reply to Dr. Prodinger. *Emerg Infect Dis* 2001; 7: 763–764.
100. Prodinger WM, Bunyaratvej P, Prachaktam R, Pavlic M. *Mycobacterium tuberculosis* isolates of Beijing genotype in Thailand. *Emerg Infect Dis* 2001; 7: 483–484.
101. Kruuner A, Hoffner SE, Sillastu H, *et al.* Spread of drug-resistant pulmonary tuberculosis in Estonia. *J Clin Microbiol* 2001; 39: 3339–3345.
102. Laserson KF, Osorio L, Sheppard JD, *et al.* Clinical and programmatic mismanagement rather than community outbreak as the cause of chronic, drug-resistant tuberculosis in Buenaventura, Colombia, 1998. *Int J Tuberc Lung Dis* 2000; 4: 673–683.
103. Portaels F, Rigouts L, Bastian I. Addressing multi-drug-resistant tuberculosis in penitentiary hospitals and in the general population of the former Soviet Union. *Int J Tuberc Lung Dis* 1999; 3: 582–588.
104. van Soolingen D, Qian L, de Haas PE, *et al.* Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia. *J Clin Microbiol* 1995; 33: 3234–3238.
105. Caminero JA, Pena MJ, Campos-Herrero MI, *et al.* Epidemiological evidence of the spread of a *Mycobacterium tuberculosis* strain of the Beijing genotype on Gran Canaria island. *Am J Respir Crit Care Med* 2001; 164: 1165–1170.
106. Anh DD, Borgdorff MW, Van LN, *et al. Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. *Emerg Infect Dis* 2000; 6: 302–305.

107. Codina G, Vidal R, Martin-Casabona N, Miravitlles M, Martin C. Multidrug-resistant tuberculosis caused by 'W'-related strains in three immunocompetent foreign-born patients. *Int J Tuberc Lung Dis* 1999; 3: 82–84.
108. Shafer RW, Singh SP, Larkin C, Small PM. Exogenous reinfection with multidrug-resistant *Mycobacterium tuberculosis* in an immunocompetent patient. *Tuber Lung Dis* 1995; 76: 575–577.
109. Godfrey-Faussett P, Sonnenberg P, Shearer SC, *et al.* Tuberculosis control and molecular epidemiology in a South African gold-mining community. *Lancet* 2000; 356: 1066–1071.
110. van Rie A, Warren R, Richardson M, *et al.* Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *N Engl J Med* 1999; 341: 1174–1179.
111. Sahadevan R, Narayanan S, Paramasivan CN, Prabhakar R, Narayanan PR. Restriction fragment length polymorphism typing of clinical isolates of *Mycobacterium tuberculosis* from patients with pulmonary tuberculosis in Madras, India, by use of direct-repeat probe. *J Clin Microbiol* 1995; 33: 3037–3039.
112. Caminero JA, Pena MJ, Campos-Herrero MI, *et al.* Exogenous reinfection with tuberculosis on a European island with a moderate incidence of disease. *Am J Respir Crit Care Med* 2001; 163: 717–720.
113. Goble M, Iseman MD, Madsen LA, Waite D, Ackerson L, Horsburgh CR Jr. Treatment of 171 patients with pulmonary tuberculosis resistant to isoniazid and rifampin. *N Engl J Med* 1993; 328: 527–532.
114. Mitchison DA, Nunn AJ. Influence of initial drug resistance on the response to short-course chemotherapy of pulmonary tuberculosis. *Am Rev Respir Dis* 1986; 133: 423–430.
115. Dye C. Will tuberculosis become resistant to all antibiotics?. *Proc R Soc Lond B* 2000; 268: 45–52.
116. Anderson RM. The pandemic of antibiotic resistance. *Nat Med* 1999; 5: 147–149.
117. Bjorkman J, Nagaev I, Berg OG, Hughes D, Andersson DI. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. *Science* 2000; 287: 1479–1482.
118. Li Z, Kelley C, Collins F, Rouse D, Morris S. Expression of *katG* in *Mycobacterium tuberculosis* is associated with its growth and persistence in mice and guinea pigs. *J Infect Dis* 1998; 177: 1030–1035.
119. Garcia-Garcia ML, Ponce de Leon A, Jimenez-Corona ME, *et al.* Clinical consequences and transmissibility of drug-resistant tuberculosis in Southern Mexico. *Arch Intern Med* 2000; 160: 630–636.